

## SISTEM REKOMENDASI JURUSAN KULIAH BAGI CALON MAHASISWA BARU MENGGUNAKAN ALGORITMA C4.5 (Studi Kasus : Universitas Muhammadiyah Riau)

Annisa Ulfa<sup>1)</sup>, Doni Winarso M.Kom<sup>2)</sup>, Edo Arribe MMSi<sup>3)</sup>.

<sup>1</sup>Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau  
email: 150402017@student.umri.ac.id

<sup>2</sup>Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau  
email: doniwinarso@umri.ac.id

<sup>3</sup>Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau  
email: edoarribe@umri.ac.id

### Abstract

*At the tertiary level, a person learns based on a particular field of expertise, so that the output of a tertiary institution is expected to be able to produce quality human resources in accordance with their area of expertise. Incorrect in choosing college majors is usually due to lack of knowledge about the existing majors. So that the selection of majors must be done as early as possible with full consideration. One way is to use a system of recommendation for the selection of majors for prospective new students based on school grades, test scores and school majors using the C4.5 algorithm. The purpose of this study is to produce a department majors recommendation system for prospective new students using the C4.5 algorithm. The output of this study is a recommendation system for prospective new students who are still confused in determining their majors. The results of the C4.5 algorithm evaluation conducted using the Confusion matrix obtained an accuracy rate of 12% and an error of 88%. This shows that the system will fail to make recommendations accordingly because the data and variables used are not supportive but if the data or variables used support the system will be able to make recommendations correctly.*

**Keywords:** C4.5, Decision Tree, Recommendations.

### Abstrak

*Pada tingkatan perguruan tinggi, seseorang belajar berdasarkan suatu bidang keahlian tertentu, sehingga output dari sebuah perguruan tinggi diharapkan mampu menghasilkan sumber daya manusia yang berkualitas sesuai dengan bidang keahliannya. Salah dalam memilih jurusan kuliah biasanya disebabkan karena kurangnya pengetahuan mengenai jurusan yang ada. Sehingga pemilihan jurusan harus dilakukan sedini mungkin dengan penuh pertimbangan. Salah satu cara dengan menggunakan sistem rekomendasi pemilihan jurusan bagi calon mahasiswa baru berdasarkan nilai sekolah, nilai tes dan jurusan sekolah menggunakan algoritma C4.5. Tujuan penelitian ini adalah menghasilkan suatu sistem rekomendasi jurusan kuliah bagi calon mahasiswa baru menggunakan algoritma C4.5. Luaran penelitian ini adalah sebuah sistem rekomendasi bagi calon mahasiswa baru yang masih kebingungan dalam menentukan jurusan kuliah. Hasil evaluasi algoritma C4.5 yang dilakukan dengan menggunakan confusion matrix didapatkan tingkat akurasi sebesar 12%. Hal ini menunjukkan bahwa sistem akan gagal melakukan rekomendasi yang sesuai karena data dan variabel yang digunakan kurang mendukung namun apabila data maupun variabel yang digunakan mendukung maka sistem akan bisa melakukan rekomendasi dengan benar.*

**Keywords:** C4.5, Decision Tree, Rekomendasi.

### PENDAHULUAN

Perguruan tinggi merupakan salah satu institusi pendidikan yang sangat penting untuk mencetak tenaga profesional. Pada tingkatan perguruan tinggi, seseorang belajar

berdasarkan suatu bidang keahlian tertentu, sehingga output dari sebuah perguruan tinggi diharapkan mampu menghasilkan sumber daya manusia yang berkualitas sesuai dengan bidang keahliannya. Berdasarkan hal diatas maka sangat penting bagi seorang mahasiswa

untuk belajar pada bidang keahlian atau jurusan yang sesuai dengan bidang minat, bakat dan kemampuannya (Niswatin, 2013)

Salah dalam memilih jurusan kuliah biasanya disebabkan karena kurangnya pengetahuan mengenai jurusan yang ada, mengikuti saran dari orangtua, ikut-ikutan teman, gengsi, atau karena tidak ada pilihan jurusan yang lain. Hal ini akan berdampak pada proses pembelajaran selama kuliah karena kurang mengerti materi-materi yang diberikan atau bahkan tidak menyukai materi perkuliahan tersebut yang akan mengakibatkan rendahnya indeks prestasi (IP) hingga terancam dikeluarkan (*DO-Drop Out*).

Sehingga pemilihan jurusan harus dilakukan sedini mungkin dengan penuh pertimbangan. Salah memilih jurusan dapat menyebabkan kerugian yang besar bagi calon Mahasiswa baru di masa depan. Oleh karena itu, calon mahasiswa baru harus mampu mengetahui kemampuan dan minat yang dimilikinya (Fitriawanti, Cholissodin, & Dewi, 2018)

Universitas Muhammadiyah Riau adalah salah satu perguruan tinggi swasta yang memiliki banyak jurusan, sehingga calon mahasiswa baru yang ingin mendaftarkan dirinya kadang merasa kebingungan dalam menentukan jurusan kuliah. Akibatnya banyak yang gagal dalam masa perkuliahan.

Kegagalan dalam masa perkuliahan ini dapat dilihat dari data mahasiswa yang di *drop out* dari tahun 2008 dimana sebanyak 124 mahasiswa, pada tahun 2009 sebanyak 171 mahasiswa, tahun 2010 sebanyak 271 mahasiswa, tahun 2011 sebanyak 263 mahasiswa, kemudian pada tahun 2012 sebanyak 256 mahasiswa, tahun 2013 sebanyak 215 mahasiswa, tahun 2014 sebanyak 204 mahasiswa dan pada tahun 2015 sebanyak 93 mahasiswa yang di *drop out*.

Salah satu cara untuk mengatasi calon mahasiswa yang salah dalam memilih jurusan ini adalah sistem pengambilan keputusan (SPK). SPK sifatnya hanya membantu para pengambil keputusan dalam menentukan pilihan dari sekian banyak alternatif berdasarkan kriteria dan pembobotan yang ditentukan, bukan sebagai pengganti pengambil keputusan dalam

memutuskan suatu hal (Winarso, Nurita, & Syahril, 2018). Pohon keputusan (*Decision Tree*) merupakan salah satu teknik pada data mining untuk menentukan klasifikasi yang dapat membantu pengambilan sebuah keputusan. Algoritma C4.5 adalah teknik pemecah masalah berdasarkan klasifikasi dengan karakteristik dimana proses penentuan berdasarkan nilai *entropy* dan nilai *gain* dari kemungkinan setiap kriteria yang kemudian akan dilanjutkan dengan perangkingan hasil keputusan. Kelebihan dari algoritma C4.5 ini memiliki tingkat akurasi yang tinggi dan dapat diterima, menghasilkan pohon keputusan yang mudah dimengerti dan dapat menangani atribut-atribut bertipe diskret dan numerik.

## METODE PENELITIAN

Metodologi penelitian ini dilakukan sebagai pedoman peneliti untuk mencapai hasil dan tujuan yang diharapkan.

Data Mining adalah proses yang menggunakan statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Data mining disisi lain adalah kegiatan meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan (Swastina, 2013).

Berikut adalah tahapan dalam data mining (Utama, Sihwi, & Doewes, 2014):

### 1. Pengumpulan Data

Pengumpulan dan pencarian data yang diperoleh dari Bagian Administrasi Akademik (BAAK) bagian Penerimaan Mahasiswa Baru di Universitas Muhammadiyah Riau. Data yang diperoleh berupa daftar penerimaan mahasiswa baru.

### 2. Seleksi Data

Seleksi data adalah pemilihan data penerimaan mahasiswa baru berdasarkan atribut atau kriteria yang telah ditentukan seperti nilai sekolah, nilai tes, dan jurusan sekolah asal.

### 3. Cleaning Data

Menghapus data-data yang tidak memenuhi atribut atau kriteria yang telah ditentukan.

### 4. Transformasi Data

Transformasi data adalah proses mengubah data kedalam bentuk atau format yang sesuai untuk diproses ke dalam perhitungan algoritma C4.5

### 5. Perhitungan Algoritma C4.5

Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai berikut: dapat mengatasi *missing value*, dapat mengatasi *continue data*, dan *prunning* (Atma & Setyanto, 2018). Tahapan dalam membuat sebuah *Decision Tree* dalam algoritma C4.5, yaitu (Sirait & Hansun, 2017) :

1. Mempersiapkan *data training*. Data training didapat dari data *history* yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dengan cara menghitung nilai *gain* dari masing-masing atribut. Nilai *gain* yang paling tinggi akan menjadi akar yang pertama. Sebelum menghitung nilai *gain* dari atribut, terlebih dahulu hitung nilai *entropy*.

$$Entropy(S) = \sum_{i=1}^n -P_i \log_2 P_i$$

Keterangan:

$S$  = himpunan kasus

$n$  = jumlah partisi  $S$

$P_i$  = proporsi  $S_i$  terhadap  $S$

3. Kemudian hitung nilai *gain* dengan menggunakan rumus :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \left| \frac{S_i}{S} \right| Entropy(S_i)$$

Keterangan:

$S$  = himpunan kasus

$A$  = fitur

$n$  = jumlah partisi atribut  $A$

$|S_i|$  = proporsi  $S_i$  terhadap  $S$

$|S|$  = jumlah kasus dalam  $S$

4. Ulangi langkah kedua dan langkah ketiga sampai *record* terpartisi.
5. Proses *decision tree* akan berhenti ketika:
  - a. Semua *record* dalam simpul  $n$  mendapat kelas yang sama
  - b. Tidak ada atribut di dalam *record* yang dipartisi lagi
  - c. Tidak ada *record* di dalam cabang yang kosong

Pada algoritma C4.5 untuk memperbaiki informasi dari *gain* dapat menggunakan *gain ratio*.

$$Gain\ Ratio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)}$$

dimana  $S$  merupakan ruang (*sample*) data yang digunakan untuk *training* dan  $A$  adalah atribut.  $Ga(S, A)$  merupakan informasi *gain* pada atribut  $A$ ,  $SplitInfo$  adalah nilai *split information* pada atribut yang didapat dari rumus sebagai berikut:

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S}$$

### 6. Pohon Keputusan (Decision Tree)

Pohon keputusan dibuat setelah melakukan perhitungan terhadap *entropy* dan *information gain* sehingga mendapatkan *information gain* tertinggi dan dilakukan secara berulang-ulang sampai atribut pohon keputusan tidak bisa melakukan proses perhitungan kembali.

### HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data penerimaan mahasiswa baru di Universitas Muhammadiyah Riau pada tahun 2018/19 yang berjumlah 1877 data mahasiswa dan 77 jenis variabel, maka ditentukan variabel yang paling sesuai dalam merekomendasi jurusan bagi mahasiswa baru adalah nilai sekolah (NS), Jurusan Sekolah (JS), dan Nilai Tes (NT).

Kemudian hitung *Entropy* dan *Information Gain* untuk setiap atribut. Untuk

Gambar 1 Perhitungan Entropy dan Information Gain

	S	S1	S2	S3	S4	S5	S6	Entropy	Gain
Total	15	1	5	2	2	4	1	2.332915	
Nilai Sekolah									0.614619
tinggi	4	1	1	1	1	0	0	2	
sedang	9	0	4	0	1	3	1	1.752715	
rendah	2	0	0	1	0	1	0	1	
Nilai Tes									0.365369
tinggi	2	0	0	1	1	0	0	1	
sedang	2	0	1	0	0	0	1	1	
rendah	11	1	4	1	1	4	0	2.31938	
Jurusan Sekolah									0.455261
tinggi	3	0	0	0	0	2	1	0.918296	
rendah	12	1	5	2	2	2	0	2.117492	

lebih jelasnya dapat dilihat pada tabel berikut.

Kemudian hitung juga *SplitInfo* dan *Gain Ratio* untuk setiap atribut. Selanjutnya pilih atribut yang memiliki *Gain Ratio* paling tinggi untuk dijadikan *root node*. Untuk lebih

Gambar 2 Perhitungan SplitInfo dan Gain Ratio

	TINGGI	SEDANG	RENDAH	SplitInfo	GainRatio
NS	-0.26667 0.266667	-0.6	0.6	-0.133333 0.133333	1.34 0.46
	-1.90689	-0.73696559	-2.90689		
	0.508504	0.442179356	0.387585		
NT	-0.133333 0.133333	-0.133333333 0.133333	-0.733333 0.733333	1.10	0.33
	-2.90689	-2.9068906	-0.44746		
	0.387585	0.387585413	0.328137		
JS	-0.2 0.2	-0.8 0.8		0.72	0.63
	-2.32193	-0.32192809			
	0.464386	0.257542476			

jelasan dapat dilihat pada tabel berikut.

Dari 100 data uji didapatkan hanya 12 data yang memiliki hasil rekomendasi sesuai. Kemudian dapat ditentukan nilai keberhasilan sistem sebagai berikut:

$$\text{Persentasi Accuracy} = \frac{3 + 2 + 7}{100} \times 100\% = 12\%$$

$$\text{Persentasi Error} = 100\% - 12\% = 88\%$$

Dari perhitungan diatas maka didapatkan hasil akurasi dari sistem sebesar 12% dengan menggunakan metode *Confusion Matrix* dan persentasi error sebesar 88%.

Dari penelitian ini menghasilkan sebuah sistem rekomendasi jurusan kuliah bagi calon mahasiswa baru dengan kecocokan rekomendasi hanya sebesar 12%. Nilai kecocokan rekomendasi ini rendah disebabkan karena variabel data yang digunakan kurang mendukung. Di Universitas Muhammadiyah Riau sendiri belum melakukan tes minat dan bakat atau tes IQ yang cukup pada calon mahasiswa baru ketika melakukan pendaftaran kuliah dan juga belum ada syarat-syarat atau ketentuan yang berlaku untuk calon mahasiswa yang ingin melakukan lintas jurusan. Sehingga calon mahasiswa yang berasal dari jurusan apapun di sekolahnya dapat memasuki jurusan apapun yang diinginkannya yang mengakibatkan rendahnya hasil kecocokan sistem rekomendasi ini.

## SIMPULAN DAN SARAN

Berdasarkan hasil implementasi dan pengujian yang telah dilakukan, maka dapat ditarik beberapa kesimpulan sebagai berikut :

1. Sistem rekomendasi jurusan bagi mahasiswa baru menggunakan algoritma C4.5 ini menghasilkan akurasi sebesar 12% dan error sebesar 80%.
2. Data yang digunakan tidak sesuai dalam merekomendasikan jurusan kuliah menggunakan algoritma C4.5. Hal ini dikarenakan data yang kurang relevan, misalnya calon mahasiswa yang jurusan asal sekolahnya ekonomi, namun ketika mendaftar di UMRI di jurusan keperawatan, calon mahasiswa tersebut tetap diterima. Sistem ini dapat digunakan apabila

datanya relevan antara input dan outputnya.

Adapun saran yang dapat diberikan untuk penelitian lanjutan yang terkait dengan penelitian ini adalah:

1. Diharapkan dapat mengembangkan sistem ini menggunakan model klasifikasi data mining lain atau dengan melakukan perbandingan metode.
2. Diharapkan dapat menganalisa ulang data dengan menggunakan atribut selain nilai sekolah, nilai tes, dan jurusan sekolah, tetapi bisa ditambahkan dengan nilai sekolah IPA, nilai sekolah IPS, nilai sekolah matematika, atau data minat bakat dan tes IQ agar mendapatkan hasil yang lebih baik.
3. Diharapkan dapat melakukan pengolahan data menggunakan *tools* lain selain Weka.

## TERIMA KASIH

Selama pelaksanaan dan pengerjaan laporan ini, penulis mendapatkan banyak pengetahuan, pengalaman, bimbingan, dukungan dan juga arahan dari semua pihak yang telah membantu hingga penulisan laporan ini dapat diselesaikan. Untuk itu, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Bapak DR. Mubarak, M.Si. selaku Rektor Universitas Muhammadiyah Riau.
2. Bapak Harun Mukhtar, M.Kom. selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Doni Winarso, M.Kom selaku Ketua Prodi Sistem Informasi dan Dosen Pembimbing I yang telah memberikan pengarahan, bimbingan, saran serta dorongan yang sangat berarti kepada penulis dalam menyusun laporan ini.
4. Bapak Edo Arribe, S.Kom., MMSI. selaku Dosen Pembimbing II yang telah memberikan bimbingan, dan arahan kepada penulis dalam menyusun laporan ini.
5. Bapak/Ibu dosen serta staf pegawai Fakultas Ilmu Komputer Universitas Muhammadiyah Riau yang telah memberikan pengetahuan yang

- bermanfaat selama saya melakukan perkuliahan.
6. Bapak/Ibu Pimpinan Universitas Muhammadiyah Riau telah memberikan izin kepada saya untuk melakukan penelitian di Universitas Muhammadiyah Riau.
  7. Orangtua serta keluarga yang terus memberikan doa dan dorongan serta semangat sehingga saya dapat menyelesaikan skripsi ini.
  8. Rekan-rekan seperjuangan Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Muhammadiyah Riau.

#### DAFTAR PUSTAKA

- Atma, Y. D., & Setyanto, A. (2018). Perbandingan Algoritma C4.5 dan K-NN dalam Identifikasi Mahasiswa Berpotensi Drop Out. *Metik Jurnal*, 2(2).
- Fitriawanti, R., Cholissodin, I., & Dewi, R. K. (2018). *Klasifikasi dan Rekomendasi Jurusan Kuliah Bagi Pelajar SMA Menggunakan Algoritme Naïve Bayes - WP*. 2(11), 4914–4922.
- Niswatin, R. K. (2013). *Sistem pendukung keputusan penempatan jurusan mahasiswa baru menggunakan metode K-Nearest Neighbor*. 1(1), 55–67.
- Sirait, G., & Hansun, S. (2017). Sistem Pendukung Keputusan Pemilihan Prodi di Universitas dengan Algoritma C4.5. *Jurnal Teknik Dan Ilmu Komputer*, 6, 357–365.
- Swastina, L. (2013). Penerapan Algoritma C4 . 5 Untuk Penentuan Jurusan Mahasiswa. *Jurnal Gema Aktualita*, 2(1).
- Utama, T. D., Sihwi, S. W., & Doewes, A. (2014). Implementasi Algoritma Iterative Dichotomiser 3 Pada Penyeleksian Program Mahasiswa Wirausaha. *JURNAL ITSMART*, 3(2), 74–83.
- Winarso, D., Nurita, F., & Syahril. (2018). Penerapan Metode Weigth Product untuk Rekomendasi Penempatan Praktek Kerja Industri. *Jurnal Resti*, 2(2), 566–571.